

Beyond election day: A knowledge Boost so we see the Forest for the Trees while standing on the ISLE with a Bag of voters

Thomas Rusch¹

¹WU Vienna, Austria

Abstract

This paper provides an introduction to nonparametric classification based on trees and learning ensembles such as Bagging, Boosting, Random Forests and Importance Sampled Learning Ensembles (ISLE). Regarding tree algorithms, I will present the classic CART approach as well as a newer approach to unbiased recursive partitioning, Conditional Inference Trees. These tree methods will provide the basic classifiers whose predictive abilities will be improved by the ensemble methods.

I will illustrate the use of these methods on a data set of the 2004 US general election. The goal is to predict a person's voting behavior in 2004 based on 74 predictor variables. They include voting records for each person from 1990 to 2004, demographics such as gender and age, the household's composition of party affiliation, the person's rank in her household or the number of attended elections when eligible to vote.

It will be shown with a 10-fold leave-group-out-cross-validation procedure that Random Forests allow most accurate predictions, closely followed by Boosting with a binomial deviance loss. Additionally, the most important variables that guide prediction will be identified.

Keywords

Ensemble learning, trees, election prediction, voter turnout, silly title.

References

- Hastie, T., Tibshirani, R. & Friedman, J. (2009). *Elements of Statistical Learning*. 3rd Edition. New York: Springer.
- Hothorn, T., Hornik, K. & Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15., 651–674.